DOCUMENT RESUME

ED 478 493 TM 035 098

AUTHOR	Boe, Erling E.; May, Henry,; Boruch, Robert F.
TITLE	Student Task Persistence in the Third International Mathematics and Science Study: A Major Source of Acheievement Differences at the National, Classroom, and Student Levels.
INSTITUTION	Pennsylvania Univ., Philadelphia. Center for Research and Evaluation in Social Policy.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC. Office of Reform Assistance and Dissemination.; National Science Foundation, Arlington, VA. Directorate for Education and Human Resources.
REPORT NO	CRESP-RR-2002-TIMSS1
PUB DATE	2002-12-00
NOTE	34p.
CONTRACT	REC-9815112;NSF-96-138;R215U980021
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Ability; *Academic Achievement; *Cross Cultural Studies; Elementary Secondary Education; International Studies; Mathematics; *Measurement Techniques; *Motivation; Sciences; Scores
IDENTIFIERS	*Task Persistence; *Third International Mathematics and Science Study

ABSTRACT

The Third International Mathematics and Science Study (TIMSS) contains no direct measures of student ability and motivation. Researchers created a new variable, Student Task Persistence (STP), that is an index of student engagement in providing answers to TIMSS questions on the background questionnaire. The purpose of this research was to investigate the possibility that cross-national differences in mathematics and science achievement can be partly explained by STP. Data came from the TIMSS Student Questionnaire and the TIMSS achievement tests that were administered to national probability samples at grades 3, 4, 7, 8, and the final year of secondary school. Multilevel analyses were undertaken to study the relationships between STP and achievement at student, classroom, and national levels. Evidence suggests that the STP variable is important for two reasons. First, it represents the only available indicator of student ability and motivation to perform the tasks required by TIMSS Second, it is one of the strongest predictors of national differences in mathematics and science achievement. At the eighth grade level, at least, national mean STP scores accounted for well over half of the variation in national mean mathematics and science achievement scores, thereby leaving less than half the national-level variation in achievement to be accounted for by other factors such as student knowledge of mathematics and science content. Two appendixes contain survey questions to be asked after the Student Background Questionnaire and the achievement tests. (Contains 5 tables and 16 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 478 493

STUDENT TASK PERSISTENCE IN THE THIRD INTERNATIONAL MATHEMATICS AND SCIENCE STUDY: A MAJOR SOURCE OF ACHIEVEMENT DIFFERENCES AT THE NATIONAL, CLASSROOM, AND STUDENT LEVELS[1]

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

E. E. Boe

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Erling E. Boe, Ph.D.
Professor of Education

Henry May, Ph.D.
Research Assistant

And

Robert F. Boruch, Ph.D.
University Trustee Professor of Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Research Report No. 2002-TIMSS1

Center for Research and Evaluation in Social Policy
Graduate School of Education
University of Pennsylvania
Philadelphia, PA 19104

Revised: December 20, 2002

TM035098

[1]Support for this research was provided by a grant (Award Number R215U980021) from the Fund for Improvement in Education (CFDA Number: 84-215U), Office of Reform Assistance and Dissemination, Office of Educational Research and Improvement, U.S. Department of Education; by a grant (Grant Number REC-9815112) from the Research on Education, Policy, and Practice Program (NSF 96-138), Division of Research, Evaluation, and Communication, Directorate for Education and Human Resources, National Science Foundation; and by the Center for Research and Evaluation in Social Policy, the Graduate School of Education of the University of Pennsylvania.

BEST COPY AVAILABLE

Executive Summary

Since the release of results from the Third International Mathematics and Science Study (TIMSS) in 1996, scholars have recognized that the central importance of TIMSS lies in its contribution to a better understanding of factors that are responsible for cross-national differences in average student achievement. Among many such factors may be differences in student ability and motivation to perform the task of completing the TIMSS achievement tests in math and science. In fact, national differences in math and science achievement scores may be determined more by differences in student test-taking ability and motivation than by differences in student knowledge of math and science content. This possibility is explored in the research reported here.

Unfortunately, TIMSS contains no direct measures of student ability and motivation. Consequently, we created a new variable, Student Task Persistence (STP), that is an index of student engagement in the task of providing answers to questions contained in the TIMSS student background questionnaire. Accordingly, the STP variable is defined operationally as the extent to which an individual student persists in providing answers to questions from the student background questionnaire, as measured by the percentage of questions answered out of all questions that were asked. We interpret this variable as an index of the ability, motivation, and willingness of students to perform the tasks of answering questions contained in background questionnaires. The purpose of this research was to investigate the possibility that cross-national differences in math and science achievement can be partly explained by this index (i.e., STP).

In the absence of past research on (or theory about) STP, we propose a three-factor framework to conceptualize possible underlying response processes that, either singly or in various combinations, could account for variation in STP. These factors identify hypothetical differences among students in terms of (a) their ability to perform the TIMSS task (e.g., to read the questions, to understand the task, and to place marks on an answer sheet), (b) their motivation to work hard at the TIMSS task (e.g., to follow the instructions, to stay focused on the task, and to try to identify the best answer), and (c) their willingness to estimate or guess correct answers (i.e., to answer questions without being certain of the correct answer). Although the STP variable is based on responses to the student background questionnaire, it is plausible that these three response processes are also involved in student performance on

TIMSS achievement tests in math and science. However, in addition to the response processes of test-taking ability, motivation, and willingness to guess, performance on such achievement tests also require specific knowledge about math and science content. By contrast, this content knowledge is not relevant to answering questions from the student background questionnaire.

Methods

The data source was the TIMSS Student Questionnaire and TIMSS achievement tests in math and science that were administered to national probability samples of students at grades 3, 4, 7, and 8, and during the final year of secondary school. The cross-national relationships between the newly constructed STP variable and students' performance on the TIMSS math and science achievement tests were examined. In addition, muitilevel analyses were undertaken to assess simultaneously the relationships between STP and achievement scores at the student, classroom, and national levels. To investigate whether observed relationships between STP and achievement scores in math and science were an artifact of anomalies in the data or confounding variables, five subsidiary analyses were made to control for outliers, composition of the sample of nations, length of the student questionnaire, order of achievement tests and student questionnaire administrations, and type of multilevel estimation method.

Results

Bivariate analyses showed statistically significant correlations at all five grade levels between national mean STP scores and mean math and science scores. Particularly strong national-level coefficients of determination (r^2) of STP with math and science scores (.62 and 60, respectively) were observed at the 8th grade level.

Of the total variability in math achievement (across all students from all TIMSS nations, as averaged for grades 3, 4, 7 and 8), about 27% was attributed to differences among nations, another 20% was attributed to differences among classrooms within nations, and the remaining 53% was attributed to differences among students within classrooms. Of these three components of variation in math achievement, the STP variable accounted for about 53% of between-nation variability, 22% of between-classroom variability within nations, and 7% of the between-student variability within classrooms. In sum, the STP variable alone explained about 22% of the total variability in math achievement across all students from all nations. Similar relationships between STP and science achievement were found at the national level.

All these relationships persisted after controlling for outliers, composition of the sample of nations, and type of multilevel estimation method. Furthermore, the length of the student questionnaire and the order of achievement test and questionnaire administration did not underlie or account for the observed STP-achievement score relationships.

Conclusion

Based on this evidence, the STP variable is important for two reasons. First, it represents the only available indicator of student ability and motivation to perform the tasks required by TIMSS. Second, it is one of the strongest predictors of national differences in math and science achievement. At the 8th grade level, at least, national mean STP scores accounted for well over half the variation in national mean math and science achievement scores, thereby leaving less than half the national-level variation in achievement to be accounted for by other factors such as student knowledge of math and science content. Thus, in seeking to understand why nations differ in average student achievement, it is necessary to recognize non-academic factors, such as student characteristics, as major sources of cross-national variability in math and science achievement.

TABLE OF CONTENTS

	Page
Executive Summary	iii
Methods	iv
Results	v
Conclusion	vi
Table of Contents	vii
List of Tables	ix
List of Figures	x
Introduction	1
Student Task Persistence (STP)	2
Research Methods	3
Data Source	3
Outcome Variables	4
Predictor Variable	4
Analysis Methods	5
Bivariate Analyses	5
Multilevel Analyses	5
Control Conditions	5
Results	7
National-Level Bivariate Analyses	7
Multilevel Analyses	7
Control for Outlying Nations	12
Control for the Composition of the Sample of Nations	12
Control for the Length of Student Questionnaire	14
Control for Order of Achievement Test and Student Questionnaire Administration	15
Control for Type of Multilevel Estimation Method	16

Page

Discussion..... 17

Significance of STP..... 17

Implications for Comparative Education Research..... 18

Test-taking Research Needed..... 19

References..... 21

Appendix A..... 23

Appendix B..... 25

List of Tables

		Page
Table 1.	Third International Mathematics and Science Study (TIMSS..... 1995): Sample sizes of students, classrooms and nations used in multi-level analyses for Grades 3, 4, 7, and 8	4
Table 2.	Correlations of National Student Task Persistence (STP)..... Percentages with National Mean Mathematics Achievement Scores: Bivariate Coefficients of Determination (r^2) at Five Grade Levels with and without Outlier Nations Trimmed	8
Table 3.	Correlations of National Student Task Persistence (STP)..... Percentages with National Mean Science Achievement Scores: Bivariate Coefficients of Determination (r^2) at Five Grade Levels with and without Outlier Nations Trimmed	8
Table 4.	Multi-Level Analysis of Mathematics Achievement Score Variance. (MSV) As a Function of Student Task Persistence (STP) in Grades 3, 4, 7 and 8	10
Table 5.	Means and Standard Deviations of National Mean STP..... Percentages for Special National Samples of Students, and Cross-National Correlations of National STP Percentages Based on Subsamples with National STP Perceptions Based on Full Samples	16

List of Figures

		Page
Figure 1.	National mean mathematics achievement scores in the 8th grade..... as a function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire).	9
Figure 2.	National mean science achievement scores in the 8th grade as a..... function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire).	9
Figure 3.	Percent distribution of total achievement score variance in..... mathematics (for all students from all nations participating in TIMSS) across three levels (student, classroom, nation) at four grade levels (3rd, 4th, 7th, 8th).	11
Figure 4.	Percent of total math achievement score variance attributable to... student task persistence (STP) at each of three levels (Student, Classroom, Nation) and at four grade levels (3rd, 4th, 7th, 8th).	11
Figure 5.	Untrimmed Full Sample: National mean mathematics achievement.. scores in the 4th grade as a function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire).	13
Figure 6.	Trimmed Outlier Nation: National mean mathematics achievement.. scores in the 4th grade as a function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire).	13

Introduction

Since the release of results from the Third International Mathematics and Science Study (TIMSS) in 1996, there has been great interest in the national rankings of academic achievement by the mass media, policy makers, professionals, and researchers in the U.S. and elsewhere. However, the sponsors of TIMSS have maintained consistently that the central importance and value of this survey do not lie in computing national rankings, but rather in the much more complex and difficult process of understanding the factors that are responsible for cross-national variation in student achievement (e.g., Peak, 1996).

To this end, TIMSS provides information about a wide array of variables that can be analyzed as potential predictors of national differences in achievement. These variables pertain to (a) student background and behavior, (b) student attitudes, beliefs, and perceptions, (c) instructional factors, (d) school factors, and (e) national demographic and economic factors.

However, TIMSS does not provide direct measures of student ability or motivation that might account for national differences in achievement. Such measures could be of two types: (a) general measures of ability (i.e., academic aptitude) and motivation to perform tasks such as learning math and science knowledge in school, or (b) specific measures of ability and motivation to perform the TIMSS test-taking task. Specific ability for the TIMSS task might include the ability to read the questions, to understand the task, and to place marks on an answer sheet. Similarly, specific motivation to work hard at the TIMSS task might include motivation to follow the instructions, to stay focused on the task, and to try to identify the best answer. It is possible that differences in national average math and science achievement scores are determined more by national differences in test-taking ability and motivation than by national differences in student knowledge of math and science content. This possibility is explored in the research reported here.

Unfortunately, no research has been reported on student ability specifically to perform the test-taking task (e.g., the ability to read and understand test questions), With respect to student motivation to perform a test-taking task, however, O'Neil, Sugrue, Abedi, Baker, and Golan (1997) investigated the effect of a monetary incentive (per test item completed correctly) on student performance on released items from a math test administered by the National Assessment of Educational Progress in 1990. The monetary incentive was shown to improve both student motivation to perform the math test and performance on the test at the 8th grade level (but not at the 12th grade

level[1]). Related research has shown direct relationships between student math self efficacy beliefs and performance on math test items (Pajares, 1996). The findings from these studies suggest that student performance on math tests is related, at least under some conditions, to their motivation to perform the test-taking task.

Schmidt, Wolfe, and Kifer (1992) investigated a different possible source of national differences in math achievement test performance. They reported national differences in the tendency of "students in some systems to omit responding when they are evidently unsure of their knowledge contrasted to the behavior of students in other systems to try to answer each question—perhaps by 'guessing'" (p. 89). They observed that the interpretation of math achievement test performance "also hinges on an understanding of the students' response processes, which are as much psychological as mathematical" (p. 88). Schmidt et al. concluded that student test scores reflect much more than the knowledge of math and science needed to answer test questions correctly.

In light of what is known from the studies reviewed above, it is plausible to speculate that national differences in student achievement are related to national differences in student test-taking ability, motivation, and willingness to estimate or guess correct answers in performing the task of completing the TIMSS achievement tests in math and science. In the absence of direct measures of student ability and motivation in TIMSS, we created a new variable from available data, henceforth referred to as Student Task Persistence (STP). We interpret STP to be an index of student ability, motivation, and willingness to estimate correct answers in performing the task of answering questions contained in the student background questionnaire. The purpose of this research was to investigate the possibility that cross-national differences in math and science achievement can be partly explained by this index of a student characteristic (i.e., STP).

Student Task Persistence (STP)

Students participating in TIMSS are asked to answer a large number of background questions in the Student Questionnaire, covering a wide range of topics such as student characteristics, learning assets in the home, out-of-school activities, attitudes about learning math and science, and the like. Although all students are expected to provide answers to all questions contained in this Questionnaire, they differ appreciably in the degree to which they actually do so. The STP variable is a measure of student engagement in the task of providing answers to questions contained in the TIMSS student background questionnaire. Thus, the STP variable is defined operationally and quantified as:

---

[1] The lack of such an incentive effect at the 12th grade level was replicated by O'Neil, Abedi, Lee, Miyoshi, and Mastergeorge (2002) with TIMSS math test items.

> The extent to which an individual student persists in giving answers to particular questions from the student background questionnaire, as measured by the percentage of questions answered out of all questions asked.[2]

In the absence of past research on (or theory about) STP, we propose a three-factor framework to conceptualize possible underlying response processes that, either singly or in various combinations, could account for variation among students in STP. These factors identify hypothetical differences among students as they respond to the TIMSS student background questionnaire:

1. Ability to perform the task (e.g., to read the questions, to understand the task, and to place marks on an answer sheet),
2. Motivation to work hard at the task (e.g., to follow the instructions, to stay focused on the task, and to try to identify the best answer), and
3. Willingness to estimate or guess correct answers (i.e., to answer questions without being certain of the correct answer).

It is plausible that these response processes (ability, motivation, and willingness to guess) are also involved in student performance on TIMSS achievement tests in math and science. However, such tests also require specific knowledge about math and science content. But this content knowledge is not relevant to answering questions from the student background questionnaire.

The main emphasis of this research was on exploring the relationship between the STP index and variation in student math and science achievement scores averaged at the national level. We also analyzed such relationships at the classroom and student levels. Thus, another novel aspect of this research was the use of a multilevel analysis method that estimated the relationship between STP and achievement scores simultaneously at three levels (student, classroom, nation), thus extending the analyses of STP to the student and classroom levels and provided for a comparison of the predictive power of STP across these three levels.

Research Methods

Data Source

TIMSS Student Questionnaire and TIMSS math and science achievement test data collected in 1995 from national probability samples of students at the 3rd, 4th, 7th, and 8th grades, and during the final year of secondary school, were used for these analyses. The number of na-

---

[2]Note that a student's STP score is computed from responses to the Student (background) Questionnaire, not from responses to the TIMSS achievement tests.

tions participating in TIMSS with reasonably complete sets of data varied from 22 to 41 (depending on grade level), while the number of students included in each national probability sample varied by nation and grade level (Table 1). See Martin and Kelly (1996, 1997) for a technical description of TIMSS.

Outcome Variables

The outcome (or dependent) variables were student achievement measures; specifically, TIMSS student achievement scores in math and in science at the 3rd, 4th, 7th, and 8th grades, and at the end of secondary school. The comprehensive (i.e., full-scale) scores for math and science (as distinguished from sub-scale scores for algebra, geometry, physics, biology, etc.) were used in these analyses.

Predictor Variable

Our focus was on a single predictor (or independent) variable (STP) defined as the percentage of particular questions each student answered out of all questions asked in the TIMSS Student Questionnaire. This questionnaire is composed of 32 main items (with an average of about five sub-items each) covering a wide range of topics. Although all students were expected to answer all applicable sub-items, the fact is that students differed in the degree to which they actually did so. Thus, STP is an index of the persistence shown by students in completing the questionnaire fully. It is referred to below as the STP score.

Table 1. *Third International Mathematics and Science Study (TIMSS 1995): Sample sizes of students, classrooms and nations used in multi-level analyses for grades 3, 4, 7, and 8*

	TIMSS Sample Sizes		
Grade	Students	Classrooms	Nations
Third	81,332	4,042	24
Fourth	94,704	4,531	26
Seventh	136,137	5,950	39
Eighth	146,883	6,442	41
End of Secondary	55,684	2,990[a]	22

Note: Data from the Third International Mathematics and Science Study (TIMSS 1995).

[a]A mixture of classrooms and schools. Twelve nations sampled students within classrooms within schools, while ten nations sampled students within schools.

Analysis Methods

Bivariate Analyses. Initial bivariate analyses evaluated relationships between national mean STP scores and national mean math (and science) achievement scores for the 3rd, 4th, 7th, and 8th grades, as well as to the final year of secondary school. In these computations the unit of analysis was nation, with the number of observations (N) determined by the number of nations providing questionnaire data. The bivariate relationships were quantified by product-moment correlation coefficients.

Multilevel Analyses. Multilevel analyses of math achievement scores[3] were performed next because this method provided estimates of the relationships between STP and achievement scores simultaneously at the three levels of student, classroom, and nation. An exponential transformation of the STP variable was performed first to improve the linearity of STP-achievement relationships at the student and classroom levels. The multilevel analyses were based on a group-centered contextual model whose parameters were estimated via the restricted maximum likelihood (REML) method of hierarchical linear modeling (HLM) (Bryk & Raudenbush, 1992). TIMSS sampling weights were used at the classroom and student levels. Nations were weighted equally. This method of analysis permits computation of components of achievement variation attributable to each of the three levels, and of the proportion of achievement variability within each level that is associated with STP.

Control Conditions. To investigate whether observed relationships between STP and achievement scores in math and science were an artifact of anomalies in the data or confounding variables, five subsidiary analyses were undertaken.

1. **Outlier Analysis**: In the cross-national bivariate analyses, an outlying nation (or nations) can distort the general relationship between variables by either exaggerating or deflating the observed correlation. Therefore, scatterplots were constructed for all bivariate relationships between national mean STP scores and mean achievement scores to identify by inspection any distinctly outlier nations. Such nations were trimmed (i.e., removed from the sample) and the bivariate relationships were recomputed. Both non-trimmed and trimmed correlations are reported.
2. **Composition of Sample of Nations**. In the cross-national bivariate analyses, a correlation between national mean STP scores and mean achievement scores observed at one grade level based on one set of nations (e.g., 4th grade with 26 nations) could be quite different than a correlation observed at a different grade level based on a larger set of nations (e.g.,

[3] Multilevel analyses were not performed with science achievement scores because the TIMSS sampling design resulted in poor matching between students and science teachers (who represented the classroom level). This precluded a classroom level analysis because the linkages needed to estimate classroom level variables for a nation were not sufficiently valid.

8th grade with 41 nations). To determine whether an observed difference between correlations at two grade levels was an artifact of the composition of the sample of nations on which they were based, such correlations were also computed using the same set of nations at both grade levels.

3. **Length of Student Questionnaire**. The number of responses students were expected to make to questions and subquestions (i.e., particular items) included in student questionnaires varied considerably by nation at all TIMSS grade levels (3rd, 4th, 7th, 8th, and the end of secondary school). This variability was highest at the 7th and 8th grade levels because nations could elect to administer one of two versions of the student questionnaire [one was labeled the "Student Questionnaire: Population 2, while the other included more questions about studying specific sciences (i.e., biology, earth sciences, chemistry, physics) and was labeled the "Student Questionnaire: Population 2(s)]. It was therefore possible that national mean student item response percentages (i.e., the national STP score) might simply decline due to student fatigue or disinterest as the number of items included in the questionnaire increased. Therefore, the mean number of items included in the student questionnaire as administered in each nation was computed, and correlated with national mean STP scores and national mean math scores to investigate whether these variables were related.
4. **Order of Achievement Test and Student Questionnaire Administration**. In the administration of TIMSS, students first completed the math and science achievement tests and next were asked to complete the Student Questionnaire (from which the item non-responses were used to compute the STP score). It is therefore possible that the observed relationships between STP and student achievement scores might be an artifact of the order in which the achievement tests and the student questionnaires were administered (i.e., the order hypothesis). Specifically, to the extent that students perform poorly on the TIMSS achievement tests, it is possible that they may tend to become discouraged and disinterested in completing diligently the subsequent questionnaire task (i.e. they will <u>not</u> try as hard to complete the many questionnaire items, thereby yielding lower STP scores). If so, variation in STP scores can be partly explained by the level of achievement test performance, instead of vice versa. Therefore, further analyses were performed to investigate this matter. If national mean STP scores are determined by the achievement test performance of individual students (i.e., the order hypothesis), then controlling for individual student achievement (and thus the degree of potential discouragement) should equalize national STP scores. Thus, for a given level of achievement across all students (e.g., students performing at the international 15th percentile on the math test), the national STP scores for a

subgroup of students at the same achievement level should vary little. On the other hand, if the national STP scores vary greatly and similarly to the national STP scores for all students, then it is clear that the level of individual achievement test performance does not account for variation in national STP scores. Furthermore, if the national STP scores for these subgroups of students are highly correlated with the original national STP scores based on all students, then controlling for individual achievement does not produce a change in national STP scores, thereby suggesting that the relevant determinants of STP exist at the national level instead of the student level.

5. **Type of Multilevel Estimation Method**. The sample of nations included in TIMSS were not a random sample of nations in the world. Under this condition, K. D. Gregory (personal communication, April 3, 2002) raised the possibility that the REML method of HLM could produce biased results and that a Bayesian method of multilevel analyses via Gibbs sampling should yield results that are not as susceptible to bias (Raudenbush, Cheong, & Fotiu, 1994). Accordingly, we used both the REML and Bayesian methods with 8th grade TIMSS data to ascertain whether the results produced by the two methods were substantially different.

Results

National-Level Bivariate Analyses

As seen in the top row of Table 2 for the full sample of nations that participated in TIMSS at each grade level, the bivariate correlations between national mean STP scores and mean math scores were statistically significant at all five grade levels. Similar results for mean science scores are shown in the top row of Table 3. Particularly strong correlations for both math and science were observed at the 7th and 8th grade levels (coefficients of determination, r^2, ranged from .52 to .62). Scatterplots of the STP-achievement score correlations for 8th grade math and science are presented in Figures 1 and 2, respectively.

Multilevel Analyses

The results of the multilevel analyses of math achievement score variance (MSV) for grades 3, 4, 7, and 8 are shown in Table 4. These analyses for grade 8 show that, of the total MSV (across all grade 8 students from all 41 TIMSS nations), about 30% was attributed to differences among nations, another 21% was attributed to differences among classrooms within nations, and the remaining 50% was attributed to differences among students within classrooms. Of these three components of MSV for grade 8, the STP variable accounted for 70% of between-nation variance, 27% of between-classroom variance within nations, and 4% of the between-student variance within classrooms. In total, the STP variable alone explained 28% of total MSV across all grade 8 students from all 41 nations.

Table 2. *Correlations of National Student Task Persistence (STP) Percentages with National Mean Mathematics Achievement Scores: Bivariate Coefficients of Determination (r^2) at Five Grade Levels with and without Outlier Nations Trimmed*

TIMSS Set of Nations	Statistic	Grade Level									
		Grade 3		Grade 4		Grade 7		Grade 8		End of Secondary	
		Full	Trimmed[a]	Full	Trimmed[a]	Full	Trimmed[a]	Full	Trimmed[a]	Full	Trimmed[a]
Full TIMSS Sample	r^2	.53	.47	.33	.29	.52	-	.62	-	.33	.00
	p <	.001	.001	.01	.01	.001	-	.001	-	.01	.79
	(n)	(24)	(23)	(26)	(25)	(39)	-	(41)	-	(22)	(21)
Cohort Nations: 4th & 8th grades	r^2			.33	.29			.57	.57		
	p <			.01	.01			.001	.001		
	(n)			(26)	(25)			(26)	(25)		
Cohort Nations: 8th grade & End of Secondary	r^2							.65	.20	.37	.00
	p <							.001	.05	.01	.97
	(n)							(21)	(20)	(21)	(20)

Note. Data from the Third International Mathematics and Science Study (TIMSS 1995), IEA.

[a]Trimmed refers to the exclusion of distinctly outlying nations from the correlational analysis.

Table 3. *Correlations of National Student Task Persistence (STP) Percentages with National Mean Science Achievement Scores: Bivariate Coefficients of Determination (r^2) at Five Grade Levels with and without Outlier Nations Trimmed*

TIMSS Set of Nations	Statistic	Grade Level									
		Grade 3		Grade 4		Grade 7		Grade 8		End of Secondary	
		Full	Trimmed[a]	Full	Trimmed[a]	Full	Trimmed[a]	Full	Trimmed[a]	Full	Trimmed[a]
Full TIMSS Sample	r^2	.51	.23	.33	.13	.55	-	.60	-	.43	.01
	p <	.001	.05	.01	.10	.001	-	.001	-	.001	.66
	(n)	(24)	(23)	(26)	(25)	(39)	-	(41)	-	(22)	(21)
Cohort Nations: 4th & 8th grades	r^2			.33	.13			.54	.57		
	p <			.01	.10			.001	.001		
	(n)			(26)	(25)			(26)	(25)		
Cohort Nations: 8th grade & End of Secondary	r^2							.62	.16	.48	.03
	p <							.001	.10	.001	.47
	(n)							(21)	(20)	(21)	(20)

Note. Data from the Third International Mathematics and Science Study (TIMSS 1995), IEA.

[a]Trimmed refers to the exclusion of distinctly outlying nations from the correlational analysis.

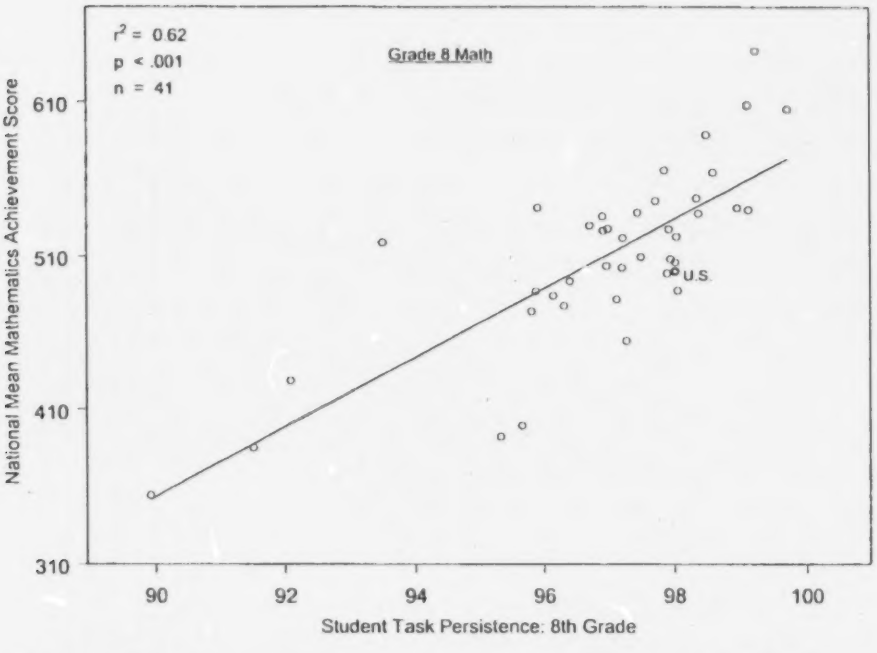

Figure 1. National mean mathematics achievement scores in the 8th grade as a function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire). Source: TIMSS

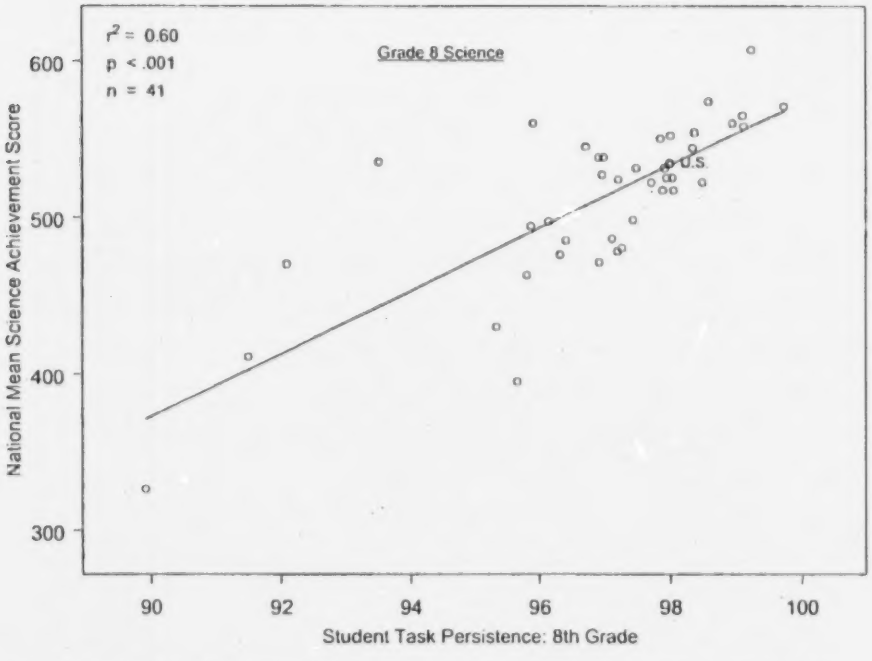

Figure 2. National mean science achievement scores in the 8th grade as a function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire). Source: TIMSS.

Table 4. *Multi-Level Analysis of Mathematics Achievement Score Variance (MSV) As a Function of Student Task Persistence (STP) in the 3rd, 4th, 7th and 8th Grades*

Grade	Analysis	Level of Analysis			
		Student	Classroom	Nation	Total
3rd	A. Percent distribution of MSV[a]	56%	21%	23%	100%
	B. Percent of MSV attributable to STP[b]	10%	19%	56%	–
	C. Percent of total MSV attributable to STP[c]	6%	4%	13%	23%
4th	A. Percent distribution of MSV[a]	54%	19%	27%	100%
	B. Percent of MSV attributable to STP[b]	8%	18%	34%	–
	C. Percent of total MSV attributable to STP[c]	4%	3%	9%	16%
7th	A. Percent distribution of MSV[a]	53%	19%	29%	100%
	B. Percent of MSV attributable to STP[b]	4%	22%	50%	--
	C. Percent of total MSV attributable to STP[c]	2%	4%	15%	21%
8th	A. Percent distribution of MSV[a]	50%	21%	30%	100%
	B. Percent of MSV attributable to STP[b]	4%	27%	70%	–
	C. Percent of total MSV attributable to STP[c]	2%	6%	20%	28%

Note: Data from the Third International Mathematics and Science Study (TIMSS 1995).

[a]In Analysis A, the total math achievement score variance (MSV) among students in all nations participating in TIMSS is distributed among three levels and sums to 100%.

[b]In Analysis B, the percent of MSV, attributable to Student Task Persistence (STP), is recorded by level. These values cannot be summed.

[c]In Analysis C, the Row A percent is multiplied by the Row B percent to obtain the percent of total MSV attributable to STP by level, and then summed across levels to obtain the percent of total MSV attributable to STP across levels.

As shown in Figure 3, the percentage distribution of MSV across the three levels (student, classroom, nation) was similar for grades 3, 4, 7, and 8. Across these four grades, there was a small decline in the percentage attributed to the student level and a corresponding small increase in the percentage attributed to the national level.

The amount of MSV explained by the STP variable by level (student, classroom, nation) is shown in Figure 4 for grades 3, 4, 7, and 8. Across these four grades, there was a considerable decline in the percentage explained at the student level and a corresponding increase in the percentage explained at the national level.

The national-level results found with the multilevel method (HLM) replicated those from the bivariate analysis. The percentage of MSV explained at the national level (as seen in Figure 4) corresponded very closely with the bivariate coefficients of determination (r^2) reported in the first row of Table 2 for the full sample of nations. This close correspondence was observed even though the multilevel analysis was based on a-transformed STP variable. Thus, the strong cross-national relationship between mean STP scores and mean math achievement scores was found by two different statistical methods at four grade levels.

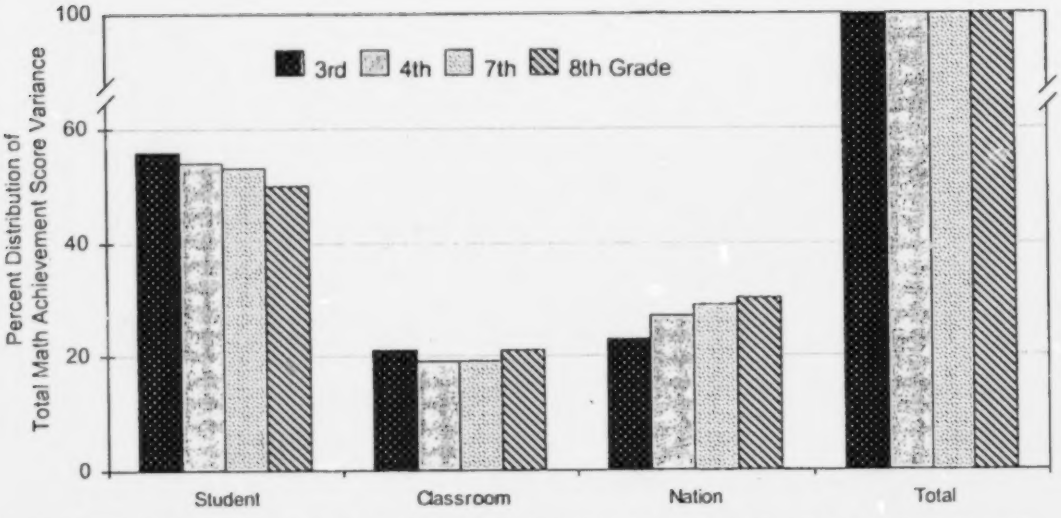

Figure 3. Percent distribution of total achievement score variance in mathematics (for all students from all nations participating in TIMSS) across three levels (student, classroom, nation) at four grade levels (3rd, 4th, 7th, 8th). Source: TIMSS.

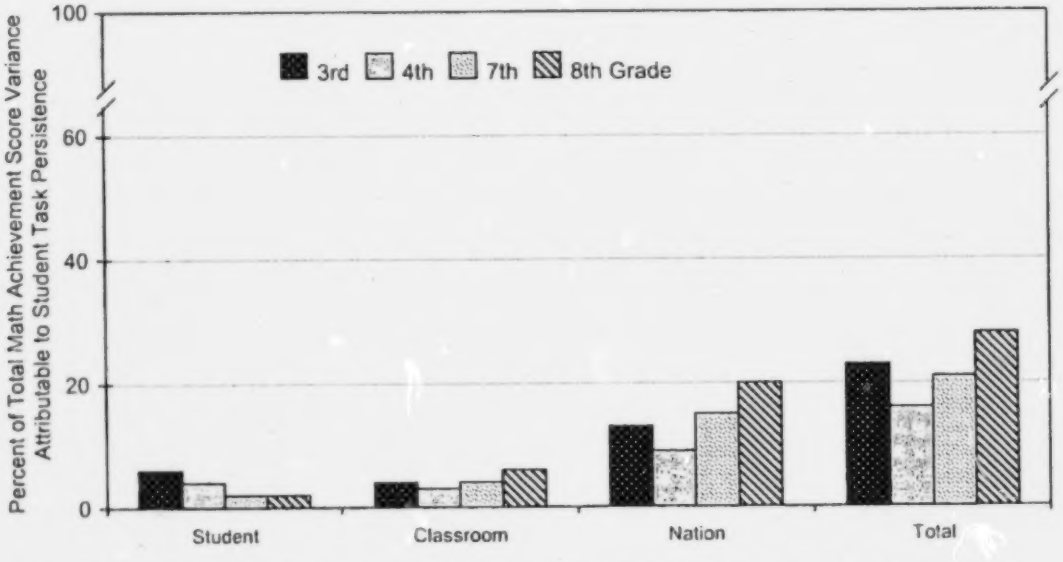

Figure 4. Percent of total math achievement score variance attributable to student task persistence (STP) at each of three levels (Student, Classroom, Nation) and at four grade levels (3rd, 4th, 7th, 8th). Source: TIMSS.

Control for Outlying Nations

Scatterplots were constructed to examine each national-level bivariate relationship based on the full sample of nations (as seen in the top rows of Tables 2 and 3) to determine if these correlations were either inflated or deflated by one or more outlying nations. It was apparent that the STP-achievement correlations at the 3rd and 4th grade levels were inflated substantially by outlying nations, as illustrated in Figure 5 for 4th grade math (based on the untrimmed full sample). When the outlier nation (South Africa, as seen in the lower left quadrant of Figure 5) was removed (i.e., trimmed), a lower correlation was obtained (see Figure 6 and the "trimmed" column of Table 2 for Grade 4). As seen in the top rows of Tables 2 and 3, this trimming procedure was performed for grade 3 and 4 correlations in both math and science.

In math, the trimming of outlier nations did not result in substantial lowering of these correlations, and correlations with both the full sample and the trimmed sample were statistically significant and substantial (trimmed r^2 of .47 and .29 at the 3rd and 4th grades, respectively). In contrast, trimming in science substantially reduced the correlations and levels of statistical significance at grades 3 and 4. For students at the end of secondary school, top rows of Tables 2 and 3 show that there was no correlation between national mean STP scores and either mean math or science scores when the one outlying nation was trimmed.

Control for the Composition of the Sample of Nations

After adjustments to the STP-achievement score correlations for outlying nations were completed, we then addressed the possible influence of variation in the sample of nations on which these correlations were based. In view of the strong STP-achievement score correlations seen for both math and science at grades 7 and 8 and the lower correlations seen at grades 3, 4, and the end of secondary school, further analyses were performed to investigate whether the lower correlations were an artifact of the different sample of nations participating in TIMSS at grade 4 (sample of 26 nations), grade 8 (sample of 41 nations), and end of secondary school (sample of 22 nations), or due to an inherent difference in the predictive power of the STP variable as a function of grade level. This was addressed by controlling for the varying composition of the sample of nations. Accordingly, analyses were repeated using only the cohort of nations that participated in TIMSS at both grades 4 and 8 (i.e., the subset of 26 nations) and only the cohort of nations that participated at both grade 8 and the end of secondary (i.e., the subset of 21 nations). Thus, any differences observed in the correlations would be related to age/grade differences rather than to sample composition.

As regards the comparison of STP-achievement score correlations between grade 4 and 8, controlling for the sample of nations and trimming outlier nations did not change substantially the size of the grade 8 correlations for either math or science (see the fourth row of Tables 2

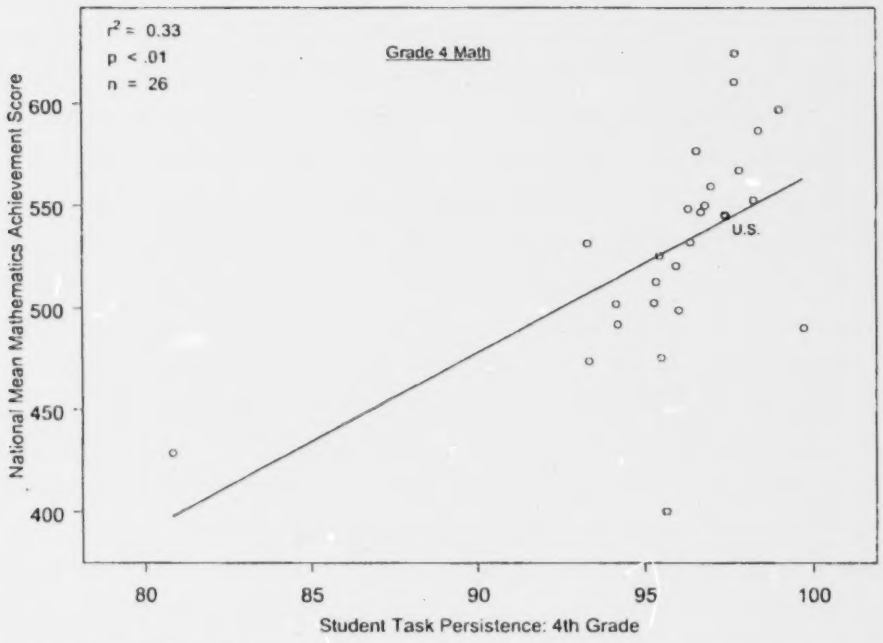

Figure 5. Untrimmed Full Sample: National mean mathematics achievement scores in the 4th grade as a function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire). Source: TIMSS.

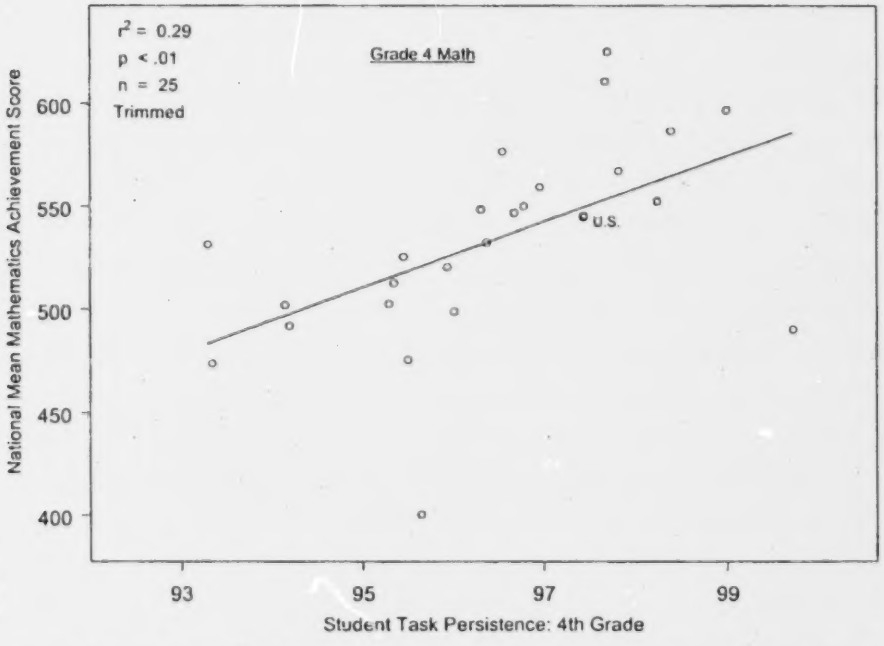

Figure 6. Trimmed Outlier Nation: National mean mathematics achievement scores in the 4th grade as a function of national Student Task Persistence (STP) percentage (i.e., for each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire). Source: TIMSS.

and 3). This demonstrates a general trend in both math and science that the relationship of STP with student achievement was somewhat larger at grade 8 than grade 4, and that this was not an artifact of sample composition.

Curiously, the opposite finding was seen when controlling for the composition of the sample of nations in comparing the STP-achievement score correlations between grade 8 and the end of secondary school. In this instance, controlling for the sample of nations and trimming outlier nations appeared to wash out these associations in both grade levels (see trimmed correlations in the seventh row of Tables 2 and 3). This finding demonstrates that the low end-of-secondary trimmed correlations (seen in the top rows of Tables 2 and 3) were mainly a product of sample composition. This is further supported by inspection of the particular nations that were excluded from the grade 8 sample of nations to make it comparable to the sample participating at the end of secondary. Specifically, among the 21 nations excluded in the math analysis were all five of the lowest scoring nations in math achievement and all four of the highest scoring nations at grade 8.[4] Thus, it is reasonable to assume that the range of national achievement scores was severely truncated at the end of secondary school, thereby reducing the probability of observing a cross-national association of math achievement with STP.[5] It is therefore likely that, had all 41 nations participated in TIMSS at the end of secondary, a substantial correlation between STP and student achievement at the end of secondary would have been observed.

Control for Length of Student Questionnaire

The mean number of particular items (i.e., questions and subquestions) included in the student questionnaires were computed for each nation participating at each grade level in TIMSS (i.e., 3rd, 4th, 7th, 8th, and end of secondary school). Bivariate correlations between mean national questionnaire items and mean math achievement scores were computed at each grade level. Similar cross-national correlations between the mean number of questionnaire items and national STP scores were computed at each grade level. None of these correlations approached statistical significance. Therefore, there is no evidence that the national-level correlations between STP and math achievement were confounded with the length of the student questionnaire.

---

[4] The five lowest scoring nations (starting with the lowest: South Africa, Columbia, Kuwait, Iran, and Portugal) and the four highest scoring nations (starting with the highest: Singapore, Korea, Japan, and Hong Kong) can be seen in the Figure 2 scatterplot for 8th grade math.

[5] Evidence for the plausibility of this assumption about cross-grade correlations at the national level comes from a correlation of average math achievement scores between 7th and 8th grades of +0.986, with an N of 39 nations.

Control for Order of Achievement Test and Student Questionnaire Administration

In the administration of TIMSS, students first completed the math and science achievement tests and next were asked to complete the Student Questionnaire (the item non-responses to which were used to compute STP scores). To control for the possibility that the observed relationships between STP and student achievement scores might be an artifact of the order in which the achievement tests and the student questionnaires were administered, we computed national mean STP scores (based on the exponentially-transformed STP variable described in the method section for "Multilevel Analyses") for students who scored between 390 and 410 on the math test (i.e., the international 15th percentile, plus or minus 10 points). If students who scored low on the math test became discouraged and disinterested in completing diligently the subsequent questionnaire task, the variability of their national mean STP scores should be much lower than those computed from the full sample of students, and their national mean STP scores should not be correlated strongly with the STP scores computed from the full sample of students. In other words, the degree of discouragement experienced by a student resulting from poor performance on the achievement test should not depend on that student's nationality. Such a dependence on nationality would suggest a national-level effect (e.g., national motivation to persist on the TIMSS task) that is different than the individual-level effect of achievement test performance analyzed here.

We also extended this logic to students who scored between 590 and 610 on the math test (i.e., the international 80th percentile, plus or minus 10 points). Such students who scored high on the math test presumably would not became discouraged and disinterested in completing diligently the subsequent questionnaire task. Therefore, the variability of their national mean STP scores should likewise be much lower than those computed from the full sample of students, and their national mean STP scores should not be correlated strongly with the STP scores computed from the full sample of students. For comparison, we also examined the STP scores of students who scored between 490 and 510 on the math test (i.e., the international 50th percentile, plus or minus 10 points).

The results of these analyses are shown in Table 5. It is apparent that the standard deviations of national mean STP scores are not substantially lower in the three truncated national samples (i.e., as drawn from the international 15th, 50th, and 80th percentiles of math scores) than in the full national sample. In addition, the national mean STP scores for the three truncated samples are all highly correlated with the national mean STP scores in the full national sample. Therefore, no evidence was found to support the hypothesis that national-level STP scores can be partly explained by the level of individual achievement test performance.

Table 5. *Means and Standard Deviations of National Mean STP Percentages for Special National Samples of 8th Grade Students in Comparison with the Full National Samples of Students, and Cross-National Correlations of National STP Percentages Based on Subsamples with National STP Perceptions Based on Full Samples*

Three TIMSS Subsamples Based on Student Math Score	National STP Percentages[a]			Cross-National Correlation (r) of Subsample STP %s with Full Sample STP %s
	N	Mean	Std Dev.	
1. International 15th Percentile (Math Scores: 390-410)	41	65.3%	10.7%	.79
2. International 50th Percentile (Math Scores: 490-510)	41	76.4%	8.4%	.82
3. International 80th Percentile (Math Scores: 590-610)	40	82.3%	8.6%	.82
Full International Sample	41	74.4%	10.9%	–

Note: Data from the Third International Mathematics and Science Study (TIMSS 1995).

[a]STP Percentages: For each nation, the percentage of specific items actually answered by students out of the total number of specific items asked in the TIMSS Student Questionnaire. These percentages were then exponentially transformed (see Method Section).

Control for Type of Multilevel Estimation Method

Potential bias in the REML method of HLM estimation attributable to the non-random sampling of nations in TIMSS was investigated using an alternative estimation method that has been shown to produce relatively unbiased variance component estimates under such circumstances (Raudenbush, et al., 1994). This Bayesian method, known as Markov Chain Monte Carlo (MCMC) using Gibbs sampling, was implemented using eighth grade data. Non-informative prior distributions[6] were defined for both the fixed and random effects. REML estimates were used as the staring values for the Gibbs sampler. After a 5,000 iteration burn-in period, 25,000 draws were sampled from the posterior distribution of parameter estimates.

The MCMC estimates of both the fixed and random effects were very close to those from the REML method. The variance partition of mathematics achievement from the MCMC method was nearly identical to that of the REML method, with each MSV component within 1% of the corresponding estimate from the alternative model. The within-nation fixed effects from the full model were also within 1% of the corresponding original estimates, and the nation-level fixed effect decreased by only 7%. The MCMC estimates of the percent of math achievement vari-

[6] See Congdon (2001) and Raudenbush & Bryk (2002) for a discussion of Bayesian modeling and prior distributions.

ance explained by the STP variable at each level were all within 1.5% of the corresponding REML estimates. Therefore, no evidence was found to suggest that the non-random sampling of nations in TIMSS produced biased estimates of parameters under the REML estimation.

Discussion

In the absence of past research and theory in comparative education about the impact of student test-taking ability and motivation on cross-national differences in mean achievement scores, the research reported here into a newly-created variable (STP) was the first to explore such associations. In bivariate analyses, we demonstrated that national mean STP scores account for a large percentage (from 13% to 62%) of the differences among nations in mean math and science scores at grades 3, 4, 7, and 8. Though we did not observe a similar relationship for students at the end of secondary school, results of our analyses also suggest that a substantial percentage of the cross-national differences in mean math and science scores would have been associated with STP if the sample of nations available for analyses at this level had been as large and variable as that for grade 8. In addition, we also demonstrated with multilevel analyses that STP accounts for about 20% of the variability among classrooms (within nations) in math achievement, and about 7% of the variability among students (within classrooms).

In demonstrating these STP-achievement score bivariate relationships at the national level, we took into account both the impact of outlier nations and the grade-level differences in composition of the sample of nations on which the analyses were based. Further analyses demonstrated that the national STP-achievement score relationships were not confounded with variation in the length of student questionnaires or the order in which the achievement tests and student questionnaires were administered.

These exploratory analyses have revealed a new and robust empirical phenomenon. It is robust in that it is observable in two subject matters (math and science). It is robust in being observable across several elementary and middle school grades, and at the national, classroom, and student levels. With replication of the relationship between STP and achievement scores across different national probability samples of students at each grade studied, and by using two statistical methods, there is strong evidence that the observed associations are reliable and have unusually high external validity. It is therefore appropriate and important to ask: "what is the significance of STP?"

Significance of STP

The STP variable is clearly is a characteristic of individual students, specifically their propensity to provide answers to background questions (as distinguished from responses to achievement test items). Under the framework for conceptualizing STP that we proposed, vari-

ability in students' propensity to provide such answers reflects differences in their (a) ability to perform the questionnaire task, (b) motivation to perform the task, and (c) willingness to make estimates when the best answer is not clear. Whatever student characteristic(s) STP might represent, obviously it is not a direct measure of knowledge about math or science concepts because the student background questionnaire (on which STP is based) does not contain questions pertaining to such knowledge.

Given this interpretation of STP and the particularly strong association of nation mean STP scores with national mean achievement scores in math and science, it is apparent that national differences in student achievement scores are based on much more than student knowledge of math and science concepts. By 8th grade, well over 50% of the variance in these national mean achievement scores can be accounted for by variation in a single variable (STP) reflecting student psychological process involved in test-taking. Thus, in seeking to understand why average math and science achievement scores vary across-nations, it should not be assumed that this variability predominantly represents national differences in subject matter knowledge. Instead, it is necessary to consider non-academic factors as major sources of cross-national variability in achievement. In analyzing the results of the Second International Mathematics Study, Schmidt et al. (1992) similarly concluded that the interpretation of math test performance is ". . . as much psychological as mathematical" (p. 88).

In making these interpretations of the main findings of these analyses, we do not imply that variation in STP (an index of item non-responses to the Student Questionnaire) somehow influences or causes different levels of math and science test scores. (A limitation of TIMSS data is that it is based on surveys, not on experimental manipulation of variables.) Instead, we conclude that the both achievement test and questionnaire performances reflect some combination of three underlying (but not directly measured) psychological processes involved in test taking that are represented in the STP variable.

Implications for Comparative Education Research

Since over 50% of national mean differences in achievement test performance at the eighth grade level are explained by STP (as shown by our results), a question can be raised about how much can be learned about the educational determinants of student achievement by comparative education research. More particularly, it raises a question about the value of using cross-sectional measures of cumulative student learning in math and science (i.e., cumulative during all the years of their past schooling) as a means to evaluate the performance of national education processes, curricula, and systems. Nations do differ substantially in average academic achievement scores in math and science, but over half of cross-national variability can be attributed to a single non-educational factor (STP). This dramatizes the importance of including stu-

dent variables in comparative education research, especially considering that results from bivariate analyses of a wide range of educational variables (e.g., curriculum breadth, instructional methods, teacher qualifications) at grade 8 have demonstrated they are not strong predictors of cross-national differences in the comprehensive measures of math and science achievement used in TIMSS (Boe, et al., 2001). Instead, Boe, May, Barkanic, and Boruch (2002) found that student variables were the strongest predictors in multivariate models of cross-national differences in math and science achievement.

Though the STP variable accounts for considerably less of the variability in mean math achievement scores between classrooms (about 20%) than it does between nations (over 50%), similar conclusions can be drawn about the importance of a single student variable in explaining much of the between-classroom variability in student achievement. If one student characteristic alone can account for 20% of between-classroom variability, it is quite possible that multiple student variables would (in multivariate models) account for much more

Apart from STP, the multilevel analyses reported here also provided evidence about the distribution of math achievement score variance (MSV, the dependent variable alone) among the student, classroom, and nation levels (see Fig. 5). Over 25% of total variability in student achievement in math (i.e., across all students from all TIMSS nations combined, and averaged for grades 3, 4, 7, and 8) is attributable to differences among nations. This supports the need for comparative education research to explain the role of cross-national differences in producing variation in achievement outcomes. If, instead, this percentage had turned out to be small, the potential value of comparative research for understanding the sources of variation in academic achievement would be much less.

Test-Taking Research Needed

Even though we have learned much from the analyses reported here about the strong relationships between STP and student achievement, no study has been made of the extent to which each of three psychological processes is involved in STP. This is due solely to the unavailability of data that represents the potential components of STP. However, measures of these components could be included in future TIMSS administrations, as well as in other international studies. Following the conceptual framework of STP offered here, additional information could be collected on the following constructs:

1. Student test- and questionnaire-taking ability: to read and understand instructions, to read and understand questions, to be able to following instructions and properly record answers on answer sheets.

2. Student test- and questionnaire-taking motivation: to work intensively by focusing attention on the task, by trying to identify the best answers, by proceeding through the task at reasonable pace without wasting time, by carefully and consistently following instructions.
3. Student willingness to make estimates or guess when the best answer is not clear.
4. Student judgments about the test and questionnaire, and their performances on these tasks.

This information could be collected through additional questionnaire items, examples of which are shown in the Appendices A and B. Because STP is such a strong predictor of national differences in math and science achievement, measures of the psychological process underlying it are of interest in understanding the nature of achievement test performance. Such measures could also be used as control variables to assess how much variation in achievement scores can be attributed specifically to student subject matter knowledge instead of to psychological processes involved in test taking. In addition, it would be desirable to investigate a possible order effect on STP by scheduling the TIMSS achievement tests first and the student questionnaire second for some students, and vice versa for an equivalent group of students.

References

Boe, E. E., Barkanic, G., Leow, C. S., May, H., Shin, S, Singleton, J. C., Zeng, G., & Boruch, R. F. (2001). Correlates of national differences in mathematics and science achievement: Evidence from TIMSS (Data Analysis Report No. 2001-DAR1). Philadelphia: University of Pennsylvania, Graduate School of Education, Center for Research and Evaluation in Social Policy.

Boe, E. E., May, H., Barkanic, G, & Boruch, R. F. (2002) Predictors of national differences in mathematics and science achievement: Data from TIMSS for the eighth grade students (Research Rep. No. 2002-TIMSS2). Philadelphia: University of Pennsylvania, Graduate School of Education, Center for Research and Evaluation in Social Policy

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park: Sage.

Congdon, P. (2001) Bayesian Statistical Modeling. New York: Wiley.

Hunt, E. (1995). The role of intelligence in modern society. American Scientist, 83, 356-368.

Martin, M. O., & Kelly, D. L. (Eds.) (1996). Third International Mathematics and Science Study Technical Report: Volume I: Design and Development. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M. O., & Kelly, D. L. (Eds.) (1997). Third International Mathematics and Science Study Technical Report: Volume II: Implementation and Analysis (Primary and Middle School Years). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

McDermott, P.A., Mordell, M., & Stoltzfus, J.C. (2001). The organization of student performance in American schools: Discipline, motivation, verbal learning, and nonverbal learning. Journal of Educational Psychology, *93*, 65-76.

Messer, D. J. (1995). Mastery motivation: Past, present and future. In R. H. MacTurk& G. A. Morgan (Eds.), Mastery motivation: Origins, conceptualizations, and applications. Advances in applied developmental psychology (Vol. 12, pp. 293-316). Norwood, NJ: Ablex.

O'Neil, H. F., Abedi, J., Lee, C., Miyoshi, J., & Mastergeorge, A. (2002) Monetary incentives for low-stakes tests. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California at Los Angeles.

O'Neil, H. F., Sugrue, B., Abedi, J., Baker, E. L., & Golan, S. (1997). Final report of experimental studies on motivation and NAEP test performance (CSE Tech. Rep. No 427). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California at Los Angeles.

Pajares, F. Self-efficacy beliefs in academic settings. Review of Educational Research. 66, 543-578.

Peak, L. (1996). Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context. Washington, DC.: National Center for Education Statistics, U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., Cheong, Y.F.,& Fotiu, R. P. (1994) Synthesizing cross-national classroom effects data: Alternative models and methods. In M. Binkley, K. Rust, & M. Winglee (Eds.) Methodological Issues in comparative Educational Studies: The Case of the IEA Reading Literacy Study (NCES 94-469). National Center for Education Statistics, U.S. Department of Education, Washington, D.C.

Schmidt, W. H., Wolfe, R. G., & Kifer, E. (1992). The identification and description of student growth in mathematics achievement. In L. Burstein (Ed.) The IEA Study of Mathematics III: Student growth and classroom processes. Oxford: Pergamon Press.

Appendix A

Potential Survey Questions to be Presented after the

Mathematics/Science Achievement Examinations

1. How many times have you taken tests with multiple-choice items?

   a. never b. once c. 2 to 10 times d. more than 10 times

2. How many times have you taken tests with as many (or more) items as this test?

   a. never b. once c. 2 to 10 times d. more than 10 times

3. Did having multiple answer choices for most of the items help or hurt your performance?

   a. helped a lot b. helped somewhat c. hurt somewhat d. hurt a lot

4. The instructions for this test were:

   a. very confusing b. somewhat confusing c. somewhat understandable d. very understandable

5. The answer sheet was easy to use.

   a. strongly agree b. agree c. disagree d. strongly disagree

6. Compared to the effort you ordinarily make to do well on tests in your math and science classes, how much effort did you put forth in performing well on this test?

   a. much more effort b. more effort c. the same effort d. less effort

7. Before taking this test, how important did you think it was that you perform well?

   a. very important b. somewhat important c. somewhat unimportant d. very unimportant

8. On this test, how many times did you answer a test question when you were not sure of the correct answer?

a. none b. 1 to 5 items c. 5 to 10 items d. more than 10 items

9. More time to finish this test would have been helpful.

a. strongly agree b. agree c. disagree d. strongly disagree

10. Compared to how well you usually do in your mathematics and science classes, how well do you think you did on this test?

a. much better b. somewhat better c. somewhat worse d. much worse

11. Compared to the tests given in your mathematics and science classes, how difficult was this test?

a. very easy b. somewhat easy c. somewhat difficult d. very difficult

12. Given your performance on this math and science test, how did you feel after completing this test?

a. discouraged b. somewhat discouraged c. somewhat encouraged d. encouraged

Appendix B

Potential Survey Questions to be Presented after the Student Background Questionnaire

1. How many times have you answered similar questionnaires?

   a. never b. once c. 2 to 10 times d. more than 10 times

2. The instructions for this questionnaire were:

   a. very confusing b. somewhat confusing c. somewhat understandable d. very understandable

3. What was the primary reason for why you did not answer some of the items on this questionnaire?

   a. didn't know the answer
   b. didn't want to give personal information
   c. didn't understand the question
   d. wanted to finish quickly
   e. too tired to pay attention to the question
   f. not enough time to finish
   g. didn't care enough to answer

4. On this questionnaire, how many times did you answer a questionnaire item when you were not sure of the best answer?

   a. none b. 1 to 5 items c. 5 to 10 items d. more than 10 items

5. More time to finish this questionnaire would have been helpful.

   a. strongly agree b. agree c. disagree d. strongly disagree

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM035098

I. DOCUMENT IDENTIFICATION:

Title: Student task persistence in the Third International Mathematics and Science Study: A major source of achievement differences at the national, classroom, and student levels.	
Author(s): Boe, Erling E; May, Henry; + Boruch, Robert F.	
Corporate Source: Center for Research and Evaluation in Social Policy University of Pennsylvania	Publication Date: December 20, 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1 ↑ [X]

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A ↑ [ ]

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B ↑ [ ]

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Erling E. Boe	Printed Name/Position/Title: Erling E. Boe, Professor of Education	
Organization/Address: Univer. of Pennsylvania 3700 Walnut St. Philadelphia PA 19104	Telephone: 215 898 5687	FAX: 215-8775999
	E-Mail Address: boe@pobox.upenn.edu	Date: 7-9-03

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)